1352.0.55.061 - Research Paper: Statistical Matching of the HES and NHS (Methodology Advisory Committee), Jun 2004

Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 06/07/2004 First Issue

Summary

About this Release

About this Release

Statistical matching is a procedure used to link two files or datasets where each record from one of the files is matched with a record from the second file that generally does not represent the same unit, but does represent a similar unit.

The constrained and unconstrained approaches to statistical matching are investigated in this paper. The issues associated with these approaches are identified and discussed. The conditional independence assumption, for example, is inherent in the procedure. Its implication for the analysis to be done using the matched dataset must be considered carefully.

While unconstrained matching gives the closest possible match between similar pairs, constrained matching has the advantage of replicating the marginal distributions in the donor file.

These traditional approaches to statistical matching are used to match two ABS datasets: the 1998-99 Household Expenditure Survey (HES) and the 2001 National Health Survey (NHS). The matching was done to explore building a base dataset for a microsimulation model of the Pharmaceutical Benefits Scheme (PBS). The main objective was to replicate the family structures of HES into the NHS.

Constrained matching, using linear programming, was found to be a better approach in synthetically creating completely enumerated families, and making sure that persons on the NHS are sensibly assigned to families using the HES family structure.

This paper is a preliminary output from a Technical Working Group comprising MD staff and the National Centre for Social and Economic Modelling (NATSEM). The former's main interest is to explore methodological issues associated with statistical matching procedures. The latter developed the microsimulation model of the PBS and relies on ABS microdatasets to create base files for the said model. It has done the preliminary statistical matching reported in this paper.